

RECEIVED

JAN 23 2004

TECH CENTER 1600/29



1600

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/357,675D

DATE: 01/20/2004

TIME: 13:24:38

Input Set : A:\TJU.ST25.txt

Output Set: N:\CRF4\01202004\I357675D.raw

3 <110> APPLICANT: Croce, Carlo  
 5 <120> TITLE OF INVENTION: Nitrilase Homologs  
 7 <130> FILE REFERENCE: TJU-2510  
 9 <140> CURRENT APPLICATION NUMBER: 09/357,675D  
 10 <141> CURRENT FILING DATE: 1999-07-20  
 12 <150> PRIOR APPLICATION NUMBER: 60/093,350  
 13 <151> PRIOR FILING DATE: 1998-07-20  
 15 <160> NUMBER OF SEQ ID NOS: 31  
 17 <170> SOFTWARE: PatentIn version 3.2  
 19 <210> SEQ ID NO: 1  
 20 <211> LENGTH: 1416  
 21 <212> TYPE: DNA  
 22 <213> ORGANISM: cDNA Sequence  
 25 <220> FEATURE:  
 26 <221> NAME/KEY: misc\_feature  
 27 <222> LOCATION: (19)..(19)  
 28 <223> OTHER INFORMATION: n=a  
 30 <400> SEQUENCE: 1

ENTERED

W--> 31 gcccaactgc tgcggcctnt ctggtctccag accgccctcc ggatcggacc ctgcgaatgg 60  
 33 ttttggctat atcttcatgt aggacctact ccctatcccg tcggccgcgg ctgggcttca 120  
 35 tcaccaggcc tcctcacaga ttctctgtccc ttctgtgtcc tggactccgg atacctcaac 180  
 37 tctcagtact ttgtgctcag ccacaggcca gagccatggc tatctcctct tcctcctgcg 240  
 39 aactgccccct ggtggtgtgt tgccaggtaa catcgacgcc agacaagcaa cagaacttta 300  
 41 aaacatgtgc tgagctggtt cgagaggctg ccagactggg tgccctgctg gctttcctgc 360  
 43 ctgaggcatt tgacttcatt gcacgggacc ctgcagagac gctacacctg tctgaaccac 420  
 45 tgggtgggaa acttttgaa gaatacacc agcttgccag ggaatgtgga ctctggctgt 480  
 47 ccttgggtgg ttccatgag cgtggccaag actgggagca gactcagaaa atctacaatt 540  
 49 gtacagtgtc gctgaacagc aaagggcgag tagtgccac ttacaggaag acacatctgt 600  
 51 gtgacgtaga gattccaggg caggggccta tgtgtgaaag caactctacc atgctgggc 660  
 53 ccagtccttga gtcacctgtc agcacaccag caggcaagat tgggtcagct gctgtctatg 720  
 55 acatcggtt ccctgaactc tctctggcat tggctcaagc tggagcagag atacttacct 780  
 57 atccttcagc ttttggatcc attacaggcc cagccactg ggaggtgttg ctgcgggccc 840  
 59 gtgctatcga aacccagtgc tatgtagtgg cagcagcaca gtgtggacgc caccatgaga 900  
 61 agagagcaag ttatggccac agcatggtgg tagacccctg gggaacagtg gtggcccgc 960  
 63 gctctgaggg gccaggcctc tgccttgccc gaatagacct caactatctg cgacagttgc 1020  
 65 gccgacacct gctgtgttc cagcaccgca ggcctgacct ctatggcaat ctgggtcacc 1080  
 67 cactgtctta agacttgact tctgtgagtt tagacctgcc cctccccccc ccacctgcc 1140  
 69 actatgagct agtgctcatg tgacctggag gcaggatcca ggcacagctc ccctcacttg 1200  
 71 gagaaccttg actctcttga tggaacacag atgggctgct tgggaaagaa accttcacct 1260  
 73 gagcttcacc tggagtcaga ctgcagtttc agaaaaggtg aattttatat agtcattgtt 1320  
 75 tatttcatgg aaactgaagt tctgtcagg gctgagcagc actggcattg aaaaatataa 1380  
 77 taatcataaa gtcaaaaaaa aaaaaaaaaa aaaaaa 1416  
 80 <210> SEQ ID NO: 2

## RAW SEQUENCE LISTING

DATE: 01/20/2004

PATENT APPLICATION: US/09/357,675D

TIME: 13:24:38

Input Set : A:\TJU.ST25.txt

Output Set: N:\CRF4\01202004\I357675D.raw

```

81 <211> LENGTH: 23
82 <212> TYPE: DNA
83 <213> ORGANISM: Homo sapiens
85 <400> SEQUENCE: 2
86 tctgaaactg cagtctgacc tca 23
89 <210> SEQ ID NO: 3
90 <211> LENGTH: 21
91 <212> TYPE: DNA
92 <213> ORGANISM: Homo sapiens
94 <400> SEQUENCE: 3
95 caggcacagc tcccctcact t 21
98 <210> SEQ ID NO: 4
99 <211> LENGTH: 20
100 <212> TYPE: DNA
101 <213> ORGANISM: Homo sapiens
104 <220> FEATURE:
105 <221> NAME/KEY: misc_feature
106 <222> LOCATION: (3)..(3)
107 <223> OTHER INFORMATION: n is a, c, g, or t
109 <220> FEATURE:
110 <221> NAME/KEY: misc_feature
111 <222> LOCATION: (6)..(6)
112 <223> OTHER INFORMATION: n is a, c, g, or t
114 <220> FEATURE:
115 <221> NAME/KEY: misc_feature
116 <222> LOCATION: (9)..(9)
117 <223> OTHER INFORMATION: n is a, c, g, or t
119 <220> FEATURE:
120 <221> NAME/KEY: misc_feature
121 <222> LOCATION: (12)..(12)
122 <223> OTHER INFORMATION: n is a, c, g, or t
124 <220> FEATURE:
125 <221> NAME/KEY: misc_feature
126 <222> LOCATION: (18)..(18)
127 <223> OTHER INFORMATION: n is a, c, g, or t
129 <400> SEQUENCE: 4
W--> 130 gtngtncng gncaygtngt 20
133 <210> SEQ ID NO: 5
134 <211> LENGTH: 26
135 <212> TYPE: DNA
136 <213> ORGANISM: Homo sapiens
139 <220> FEATURE:
140 <221> NAME/KEY: misc_feature
141 <222> LOCATION: (6)..(6)
142 <223> OTHER INFORMATION: n is a, c, g, or t
144 <220> FEATURE:
145 <221> NAME/KEY: misc_feature
146 <222> LOCATION: (12)..(12)
147 <223> OTHER INFORMATION: y=c or t

```

## RAW SEQUENCE LISTING

DATE: 01/20/2004

PATENT APPLICATION: US/09/357,675D

TIME: 13:24:38

Input Set : A:\TJU.ST25.txt

Output Set: N:\CRF4\01202004\I357675D.raw

```

149 <220> FEATURE:
150 <221> NAME/KEY: misc_feature
151 <222> LOCATION: (15)..(15)
152 <223> OTHER INFORMATION: n is a, c, g, or t
154 <220> FEATURE:
155 <221> NAME/KEY: misc_feature
156 <222> LOCATION: (18)..(18)
157 <223> OTHER INFORMATION: n is a, c, g, or t
159 <220> FEATURE:
160 <221> NAME/KEY: misc_feature
161 <222> LOCATION: (21)..(21)
162 <223> OTHER INFORMATION: y= c or t
164 <220> FEATURE:
165 <221> NAME/KEY: misc_feature
166 <222> LOCATION: (24)..(24)
167 <223> OTHER INFORMATION: n = a,c,g, or t
169 <400> SEQUENCE: 5
W--> 170 acrtgmacrt gyttnacngt ytgngc 26
173 <210> SEQ ID NO: 6
174 <211> LENGTH: 21
175 <212> TYPE: DNA
176 <213> ORGANISM: Drosophila melanogaster
178 <400> SEQUENCE: 6
179 ggcctttgt ggcctcgact g 21
182 <210> SEQ ID NO: 7
183 <211> LENGTH: 21
184 <212> TYPE: DNA
185 <213> ORGANISM: Drosophila melanogaster
187 <400> SEQUENCE: 7
188 cgggtggcgga agttgtctgg t 21
191 <210> SEQ ID NO: 8
192 <211> LENGTH: 20
193 <212> TYPE: DNA
194 <213> ORGANISM: Caenorhabditis elegans
196 <400> SEQUENCE: 8
197 gtggcggctg ctcaaactgg 20
200 <210> SEQ ID NO: 9
201 <211> LENGTH: 21
202 <212> TYPE: DNA
203 <213> ORGANISM: Caenorhabditis elegans
205 <400> SEQUENCE: 9
206 tcgcgacgat gaacaagtcg g 21
209 <210> SEQ ID NO: 10
210 <211> LENGTH: 19
211 <212> TYPE: DNA
212 <213> ORGANISM: Homo sapiens
214 <400> SEQUENCE: 10
215 gccctccgga tcggacctt 19
218 <210> SEQ ID NO: 11

```

## RAW SEQUENCE LISTING

DATE: 01/20/2004

PATENT APPLICATION: US/09/357,675D

TIME: 13:24:38

Input Set : A:\TJU.ST25.txt

Output Set: N:\CRF4\01202004\I357675D.raw

```

219 <211> LENGTH: 20
220 <212> TYPE: DNA
221 <213> ORGANISM: Homo sapiens
223 <400> SEQUENCE: 11
224 gacctactcc ctatcccgtc 20
227 <210> SEQ ID NO: 12
228 <211> LENGTH: 21
229 <212> TYPE: DNA
230 <213> ORGANISM: Homo sapiens
232 <400> SEQUENCE: 12
233 gctgcgaagt gcacagctaa g 21
236 <210> SEQ ID NO: 13
237 <211> LENGTH: 24
238 <212> TYPE: DNA
239 <213> ORGANISM: Homo sapiens
241 <400> SEQUENCE: 13
242 aaactgaagc ctctttcctc tgac 24
245 <210> SEQ ID NO: 14
246 <211> LENGTH: 20
247 <212> TYPE: DNA
248 <213> ORGANISM: Homo sapiens
250 <400> SEQUENCE: 14
251 tgggcttcat caccaggcct 20
254 <210> SEQ ID NO: 15
255 <211> LENGTH: 22
256 <212> TYPE: DNA
257 <213> ORGANISM: Homo sapiens
259 <400> SEQUENCE: 15
260 ctgggctgag cacaaagtac tg 22
263 <210> SEQ ID NO: 16
264 <211> LENGTH: 21
265 <212> TYPE: DNA
266 <213> ORGANISM: Homo sapiens
268 <400> SEQUENCE: 16
269 gcttgctctgg cgtcgatgtt a 21
272 <210> SEQ ID NO: 17
273 <211> LENGTH: 36
274 <212> TYPE: DNA
275 <213> ORGANISM: Homo sapiens
277 <400> SEQUENCE: 17
278 tgacgtcgac atatgtcaac tctagttaat accacg 36
281 <210> SEQ ID NO: 18
282 <211> LENGTH: 25
283 <212> TYPE: DNA
284 <213> ORGANISM: Homo sapiens
286 <400> SEQUENCE: 18
287 tgggtacctc gactagctta tgtcc 25
290 <210> SEQ ID NO: 19
291 <211> LENGTH: 147

```

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/357,675D

DATE: 01/20/2004  
 TIME: 13:24:38

Input Set : A:\TJU.ST25.txt  
 Output Set: N:\CRF4\01202004\I357675D.raw

```

292 <212> TYPE: PRT
293 <213> ORGANISM: Homo sapiens
296 <220> FEATURE:
297 <221> NAME/KEY: misc_feature
298 <222> LOCATION: (82)..(82)
299 <223> OTHER INFORMATION: Xaa is an unknown amino acid
301 <400> SEQUENCE: 19
303 Met Ser Phe Arg Phe Gly Gln His Leu Ile Lys Pro Ser Val Val Phe
304 1 5 10 15
307 Leu Lys Thr Glu Leu Ser Phe Ala Leu Val Asn Arg Lys Pro Val Val
308 20 25 30
311 Pro Gly His Val Leu Val Cys Pro Leu Arg Pro Val Glu Arg Phe His
312 35 40 45
315 Asp Leu Arg Pro Asp Glu Val Ala Asp Leu Phe Gln Thr Thr Gln Arg
316 50 55 60
319 Val Gly Thr Val Val Glu Lys His Phe His Gly Thr Ser Leu Thr Phe
320 65 70 75 80
W--> 323 Ser Xaa Gln Asp Gly Pro Glu Ala Gly Gln Thr Val Lys His Val His
324 85 90 95
327 Val His Val Leu Pro Arg Lys Ala Gly Asp Phe His Arg Asn Asp Ser
328 100 105 110
331 Ile Tyr Glu Glu Leu Gln Lys His Asp Lys Glu Asp Phe Pro Ala Ser
332 115 120 125
335 Trp Arg Ser Glu Glu Glu Glu Ala Ala Glu Ala Ala Leu Arg Val
336 130 135 140
339 Tyr Phe Gln
340 145
343 <210> SEQ ID NO: 20
344 <211> LENGTH: 150
345 <212> TYPE: PRT
346 <213> ORGANISM: murine
348 <400> SEQUENCE: 20
350 Met Ser Phe Arg Phe Gly Gln His Leu Ile Lys Pro Ser Val Val Phe
351 1 5 10 15
354 Leu Lys Thr Glu Leu Ser Phe Ala Leu Val Asn Arg Lys Pro Val Val
355 20 25 30
358 Pro Gly His Val Leu Val Cys Pro Leu Arg Pro Val Glu Arg Phe Arg
359 35 40 45
362 Asp Leu His Pro Asp Glu Val Ala Asp Leu Phe Gln Val Thr Gln Arg
363 50 55 60
366 Val Gly Thr Val Val Glu Lys His Phe Gln Gly Thr Ser Ile Thr Phe
367 65 70 75 80
370 Ser Met Gln Asp Gly Pro Glu Ala Gly Gln Thr Val Lys His Val His
371 85 90 95
374 Val His Val Leu Pro Arg Lys Ala Gly Asp Phe Pro Arg Asn Asp Asn
375 100 105 110
378 Ile Tyr Asp Glu Leu Gln Lys His Asp Arg Glu Glu Glu Asp Ser Pro
379 115 120 125
382 Ala Phe Trp Arg Ser Glu Lys Glu Met Ala Ala Glu Ala Glu Ala Leu

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/357,675D

DATE: 01/20/2004  
TIME: 13:24:39

Input Set : A:\TJU.ST25.txt  
Output Set: N:\CRF4\01202004\I357675D.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:1; N Pos. 19  
Seq#:4; N Pos. 3,6,9,12,18  
Seq#:5; N Pos. 6,15,18,24  
Seq#:19; Xaa Pos. 82  
Seq#:25; Xaa Pos. 6

**VERIFICATION SUMMARY**

DATE: 01/20/2004

PATENT APPLICATION: US/09/357,675D

TIME: 13:24:39

Input Set : A:\TJU.ST25.txt

Output Set: N:\CRF4\01202004\I357675D.raw

L:31 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1 after pos.:0  
L:130 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:0  
L:170 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:0  
L:323 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:19 after pos.:80  
L:827 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:25 after pos.:0